Journal of Memory and Language 105 (2019) 60–75

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Subjecthood and linear order in linguistic encoding: Evidence from the real-time production of *wh*-questions in English and Mandarin Chinese

Monica L. Do[a,b,*], Elsi Kaiser[a]

[a] University of Southern California, 3601 Watt Way, Grace Ford Salvatori rm 301, Los Angeles, CA 90089, United States
[b] Department of Psychology, University of Pennsylvania, 425 S. University Ave, Stephen A. Levin Bldg., Philadelphia, PA 19104-6241, United States

ARTICLE INFO

Keywords:
Linguistic encoding
Language production
Visual-world eye-tracking
wh-Questions
Word order
Mandarin Chinese

ABSTRACT

We use visual world eye-tracking to provide a first look into the real-time production of an under-researched but communicatively crucial construction – *wh*-questions. We investigate whether the transition from abstract message to highly-structured utterances (linguistic encoding) is driven by linear order (positional processing) or subjecthood assignment (functional processing). Experiment 1 decouples positional and functional processes by comparing production of English declaratives versus object *wh*-questions (‘Which nurses did the maids tickle?’). Experiment 2 compares the production of declaratives versus object *wh*-questions in Mandarin Chinese to investigate potential information-focus effects on linguistic encoding and tests whether Experiment 1’s findings could be due to focus. Experiment 1 found that even though the articulation of a sentence is necessarily linear, speakers do not necessarily encode sentences in accordance with the linear order in which the words are uttered. Experiment 2 suggests that information-focus does not guide speakers' eye-movements during linguistic encoding.

Introduction

It is generally understood that the entire process of sentence production – from the conceptualization of an idea (message formulation), to selecting the lexical- and sentence-level representations for those concepts (linguistic encoding), to morphological and phonological assembly (phonological encoding), to physical articulation – happens in stages (e.g., Levelt, 1989; Griffin & Ferreira, 2006; Ferreira, 2010; Konopka & Brown-Schmidt, 2014). Because these stages can proceed in parallel, speakers do not have to wait until all the words of a sentence are fully planned before they start speaking. Instead, speakers plan only some aspects of their utterances before they start articulating their sentences, while the rest is done 'on-the-fly'. In other words, production, like comprehension, proceeds incrementally (e.g., Levelt, 1989; Schriefers, Teruel, & Meinshausen, 1998; Ferreira & Swets, 2002; Brown-Schmidt & Konopka, 2008).

The first phase of production, message formulation, is generally understood to be purely propositional; that is, the ‘stuff’ of messages consists of pre-linguistic, unordered concepts that often include information about the entities involved in an event, the relationships between those entities, and the type of event a speaker intends to communicate about (Konopka & Brown-Schmidt, 2014; Levelt, 1989). But, what happens after speakers have ‘gotten the gist’ of what it is they want to say?

Translating an abstract thought into a linguistic representation requires that each concept in a speaker’s pre-linguistic representation be – at some point – assigned both (i) a positional slot in accordance with the linear word order of the speaker’s language, as well as (ii) a grammatical function (e.g. subject, object, etc.) that indexes that concepts’ role in the syntactic structure of the sentence. These two components (positional processing and functional processing, respectively) operate at the interface of thought and language, at a level of production known as *linguistic encoding*. But, how these processes are coordinated and unfold in real-time is not yet well-understood.

On the one hand, given that the end product of sentence production appears – at least superficially – to be linear, an intuitive possibility is that linguistic encoding is ultimately driven by **positional processing:** Speakers encode their utterances starting with the most conceptually salient element in a message; they make that element the *linearly-initial* element of the sentence; and build the syntactic structure of the utterance from that initially-selected concept (e.g. Tomlin, 1995, 1997; van Nice & Dietrich, 2003; Gleitman, January, Nappa, & Trueswell, 2007; Iwasaki, Vinson, Vigliocco, Watanabe, & Arciuri, 2008; Brown-Schmidt & Konopka, 2008; Myachykov & Tomlin, 2008; Myachykov, Thompson, Scheepers, & Garrod, 2011, 2012; see also Ferreira, 1996 for related discussion).

Because a positionally-driven account of linguistic encoding is rooted in conceptual salience, this account predicts a tight correlation

* Corresponding author.

https://doi.org/10.1016/j.jml.2018.11.001
Received 31 October 2017; Received in revised form 9 November 2018
Available online 07 December 2018

between salience and order of mention, such that the most salient element in a message should also be mentioned first. Gleitman et al. (2007) tested this prediction by using the attention capture paradigm to manipulate visual salience and found precisely this: The most visually-salient element of a to-be-described image was assigned to the sentence-initial position. Later work extended these findings by showing that in addition to visual cues, salience in production can be determined by a number of factors acting in concert (e.g. Kuchinsky, 2009; Kuchinsky, Bock, & Irwin, 2011; Konopka, 2012) – including codeability/image-ability (e.g. Bock & Warren, 1985; Bock, 1986), discourse status (e.g. Arnold, Wasow, Losongco, & Ginstrom, 2000; Bock, 1977; Ferreira & Yoshita, 2003; Prat-Sala & Branigan, 2000), and animacy (e.g. Bock, Loebell, & Morey, 1992; Altman & Kemper, 2006; Christianson & Ferreira, 2005, Tanaka, Branigan, McLean, & Pickering, 2011). Under a *positionally-driven account of linguistic encoding,* the selection of the linearly-initial concept is the starting point for the process of encoding, while grammatical function assignment plays a subsequent role in linguistic encoding.

On the other hand, a large body of linguistic work has shown that the underlying syntactic structure of a linguistic representation is fundamentally *hierarchical*, even if it may seem purely linear on the surface. Thus, a different possibility is that linguistic encoding is ultimately driven by **functional processes** (e.g. Griffin & Bock, 2000; Bock, Irwin, Davidson, & Levelt, 2003; Bock, Irwin, & Davidson, 2004; Lee, Brown-Schmidt, & Watson, 2013; Momma, Slevc, & Phillips, 2016): Speakers encode their utterances starting with the construction of a syntactic frame and then assign a concept to become the *structurally-initial* element of the utterance. In English, Mandarin Chinese and many other languages, this element is the syntactic subject of the sentence.[1] Under *a syntactically-driven account of linguistic encoding*, the first concept to be encoded is the one that speakers want to make the subject of the sentence.

Evidence for syntactically-driven accounts comes from speakers' eye-movements during the planning of active versus passive sentences (Griffin & Bock, 2000). Consider an active sentence such '*The mailman is chasing the dog.*' and a passive sentence such as '*The mailman is being chased by the dog.*' Here, the syntactic subject of both sentences is the same – '*the mailman*' – but the semantic agent (e.g. 'doer') is different. In actives, the subject is the semantic agent, but in passives, the subject is the semantic patient, or the entity that the action is done to. In an eye-tracking-during-picture-description task, Griffin and Bock (2000) found that (i) there is no predictable relationship between the most visually salient aspect of the to-be-described image and the element that speakers uttered first and that (ii) speakers encode the syntactic subject of the sentence first (as indicated by their eye-movements), regardless of whether the subject is the comparatively more salient semantic agent or the less salient patient.

In other words, Griffin and Bock's work shows that linguistic encoding begins with *the syntactic subject*, not with the conceptually most accessible element of the message. In addition, because active-passive pairs provide an opportunity to tease apart the role of agenthood and the role of subjecthood in production, this work also provides evidence that syntactic subjecthood plays a stronger role in determining the starting point for linguistic encoding than does semantic agenthood. Thus, unlike positionally-driven models of encoding, where grammatical roles such as subject and object *follow from* the selection of a linearly-initial concept, functionally-driven models posit that syntactic structure provides the starting point for linguistic encoding (the subject). Thus, under this view, it is encoding of the grammatical subject into a *hierarchically*-organized syntactic structure that *mediates* the transition from abstract, unordered concepts to linearly-organized sentences.

In relatively strict word order languages like English, the distinction between positionally- versus functionally-driven models of linguistic encoding may be difficult to appreciate because the hierarchical syntactic structure and surface linear word order of an utterance are often isomorphic. The linearly-initial element is virtually always the same as the structurally-initial grammatical role – namely the subject. But, in many flexible word order languages (e.g. Russian, Finnish, Hungarian, and Korean), where the subject of a sentence can appear sentence-medially (OSV, VSO) or even sentence-finally (OVS, VOS), this is not the case. In these languages, the tension between functional and positional processes is clearer and, crucially, a model of linguistic encoding that begins with the linearly initial argument makes fundamentally different predictions than one that begins with the syntactic subject.

Indeed, it is precisely for this reason that some prior work has sought to investigate the starting point of linguistic encoding using flexible word order languages. For instance, prior work (see Myachykov, Thompson, Scheepers, & Garrod, 2011 for review) has investigated production in Russian (Myachykov & Tomlin, 2008) and Finnish (Myachykov, Garrod, & Scheepers, 2010) – two canonically SVO languages that also allow Object-Subject-Verb (OSV) order (see also Hwang & Kaiser, 2014 for similar work in Korean). However, these studies have yielded different patterns across language, and it thus remains unclear how generalizable those results are. More importantly, these studies raise an additional issue with respect to the role of discourse-pragmatic factors in linguistic encoding, many of which are not fully understood cross-linguistically (e.g. Neeleman & Reinhart, 1997; Sekerina, 1997; Kaiser & Trueswell, 2004). Non-canonical word orders in Finnish, like other flexible-word order languages, are commonly analyzed as being triggered (or licensed) by certain discourse contexts; when non-canonical word orders are presented without an appropriate discourse, they are known to cause processing difficulties (Kaiser & Trueswell, 2004). Thus, investigating the *production* of such word orders in out-of-context configurations without an appropriate discourse context complicates the interpretation of the data. It does not seem implausible that linguistic encoding would also be sensitive to discourse-pragmatic effects, but few (but see Ganushchak, Konopka, & Chen, 2014) have investigated these issues systematically.

Seeking another way to separate functional from positional processes, other work has attempted to disentangle a subject-first approach (as predicted by functional accounts) from a linear order approach (as predicted by positional accounts) by looking at languages like Tzeltal (Norcliffe, Konopka, Brown, & Levinson, 2015) and Tagalog (Sauppe, Norcliffe, Konopka, van Valin, & Levinson, 2013), both of which have *verb*-initial word orders. However, the verb in these languages morphologically encodes properties of the subject/object of transitive sentences. As a consequence, even in these languages, the subject could be 'invisibly' sentence-initial from a sentence-planning perspective. Thus, this work does not enable us to determine whether the starting point of linguistic encoding in these languages is "driven not only by the verb's placement, but also by its complex morphology" (Norcliffe et al., 2015, p. 18; see also Sauppe, Norcliffe, Konopka, van Valin, & Levinson, 2013 for a similar discussion).

In sum, although the process of linguistic encoding reflects the fundamental transition from thought to language, relatively little is known about how it unfolds in real time. Prior work with English declarative sentences renders it hard to distinguish between syntactically-driven and positionally-driven accounts because linear word order and subjecthood in English are largely confounded. Prior work in other languages has yielded divergent results, whose interpretation is also complicated by morphological and discourse-level factors.

Even beyond the question of how central a role syntactic

[1] In an overwhelming majority of the world's languages, this structurally-initial element is the syntactic subject (e.g. Dryer, 2008; Greenberg, 1966). We discuss the case of verb-initial languages below.

structure plays in the transition from non-linguistic to linguistic representation, a better understanding of whether linguistic encoding is positionally-driven or syntactically-driven taps into a fundamental question of how different types of representations across levels of production are coordinated by the language processing system. In particular, how does the language processing system integrate (i) salience-driven information at the conceptual level of representation on one hand and (ii) syntactically-structured information on the other at the onset of linguistic encoding where these two levels of representation meet?

Independently, understanding the way that the cognitive system transitions from abstract pre-linguistic representations to highly structured syntactic representations requires an understanding of how speakers – at any point in production – weigh two competing constraints: They have to balance the need to create a sufficiently interpretable linguistic representation against the need to maintain fluency during their utterance. If speakers prioritize the construction of a syntactic frame, grammatical function assignment may ultimately drive the process of linguistic encoding. But, if speakers prioritize parsimony during production, then constraints on grammatical function assignment may be de-prioritized; given that the end product of language production is a linear string, positional processing may ultimately drive linguistic encoding. Thus, a better understanding of the relation between these two types of processes taps into fundamental questions relevant not only for models of language production but also for our understanding of how human cognitive systems handle potential conflicting pressures.

To better understand whether the transition from pre-linguistic, conceptual representations to highly structured syntactic representations is guided by the surface-level output of a sentence (i.e., via positional processing) or the underlying grammatical structures (i.e., via functional processing), we investigate how encoding unfolds when the predictions of positional and functional processes diverge. Unlike prior work, we largely sidestep concerns stemming from discourse-pragmatic and morphological factors by contrasting the production of English declaratives (1a) and *object wh*-questions (1b). We focus on the production of object *wh*-questions for two reasons.

(1a) Declarative:	**The nurses**	tickled	<u>the maids</u>.	
	Subject	Verb	Object	
(1b) Object *wh*-Question:	<u>Which maids</u>	did	**the nurses**	tickle?
	Object		Subject	Verb

First, in contrast to declaratives, where the subject precedes the object, the subject of English object *wh*-questions is not the linearly initial element of the sentence.[2] Instead, English object *wh*-questions require the object to appear before the subject. Object *wh*-questions consequently provide an ideal test-case for investigating how linguistic encoding enfolds in precisely those cases where we expect the tension between functional and positional constraints to emerge: Because the linearly-initial element (‘*which maids*’) is the syntactic object, not the subject, we can ask whether speakers will encode the subject ‘*the nurses*’ first (as syntactically-driven, functional approaches predict) or whether they will encode the object ‘*which maids*’ first (as positionally-driven approaches would predict).

Because the output of the language production process is ultimately linear, the most ‘efficient’ approach to encoding would appear to be one that begins encoding with the concept that the speaker knows must be uttered first – the object *wh*-phrase. Indeed, from a positional point-of-view, it seems ‘intuitive’ to first encode the object *wh*-phrase, because it consistently precedes the subject. But, given that other work (Griffin & Bock, 2000; Lee et al., 2013; Momma et al., 2016) has shown that speakers *can* engage in syntactically-driven encoding, it may be overly-simplistic to take the positionally-driven model of encoding as a given. Thus, an important question is under what conditions *can* we find evidence of syntactically driven planning? If we find evidence for functional processing – i.e. that speakers encode the *subject* first even though they do not mention it until after the object – this would be strong evidence for the primacy of grammatical functions during linguistic encoding, even when the deck is effectively stacked against the syntactically-driven starting point to encoding.

The second reason that we focus on the production of *wh*-questions is because they are a prevalent part of naturalistic communication. The ability to ask and answer questions is a key aspect of human interaction. For example, *wh*-questions have been shown to be critical for children’s vocabulary development and verbal reasoning skills (e.g. Hoff-Ginsberg, 1985; Rowland, Pine, Lieven, & Theakston, 2003; Rowe, Leech, & Cabrera, 2016). In addition, a large body of cross-linguistic research on *comprehension* has already shown *wh*-questions to (i) engage different processing mechanisms than declaratives and (ii) yield different processing costs than declaratives (e.g. Stowe, 1986; Frazier & Flores d’Arcais, 1989; Gibson, 2000; Sussman & Sedivy, 2003; Aoshima, Phillips, & Weinberg, 2004; Wagers, 2013). Yet, the majority of work in language production has been on production of declarative sentences. The production of questions has been largely overlooked in the domain of production. Strikingly, we know of no prior work that has investigated the production of questions using on-line methodologies; nor are we aware of any prior work investigating the production of long-distance dependencies of this type – namely, how speakers engage in incremental planning of a syntactic structure that is ‘non-incremental’ in the sense that the object *wh*-phrase is pronounced at the start of the sentence but interpreted as the object of the verb. Thus, extending work in language production to *wh*-questions allows us to investigate a novel structure that bears on an empirically- and theoretically-meaningful issue in the literature.

The current study

Experiment 1: Declaratives and wh-questions in English

In Experiment 1 we use a production-during-eye-tracking task combined with a picture-word interference manipulation to investigate whether the starting point of linguistic encoding is driven by (i) assignment of pre-linguistic concepts to syntactic roles (functional processing), or (ii) direct assignment of those concepts to a position in the sentence (positional processing) onto which syntactic roles are superimposed. We contrast the real-time production of canonical English declaratives versus object *wh*-questions, with the latter being our critical condition of interest because it separates the subject from the linearly initial slot in the sentence – enabling us to tease apart the predictions of the two accounts.

If linguistic encoding is primarily driven by direct assignment of an abstract, pre-linguistic concept to the linearly-initial position in the sentence, then the order in which speakers encode their utterances should reflect the linear word order of the sentence: When preparing to encode an object *wh*-question, speakers should encode the object before encoding the subject. By contrast, if linguistic encoding is primarily driven by grammatical role assignment then speakers should encode the subject before the object irrespective of the linear word order of the sentence. That is, speakers should encode the syntactic subject first, even when the subject is not the linearly initial element in the sentence.

[2] Under some circumstances, English does not require the object of object *wh*-questions to be fronted. This is the case of echo questions like ‘The nurses tickled *which maids*?’, which occur in specific contexts (e.g. if one person did not hear what another person said.) We do not test such contexts in this work. Indeed, the felicity of non-canonical word orders in English (as in other languages) is highly constrained by discourse context (Birner & Ward, 1998).

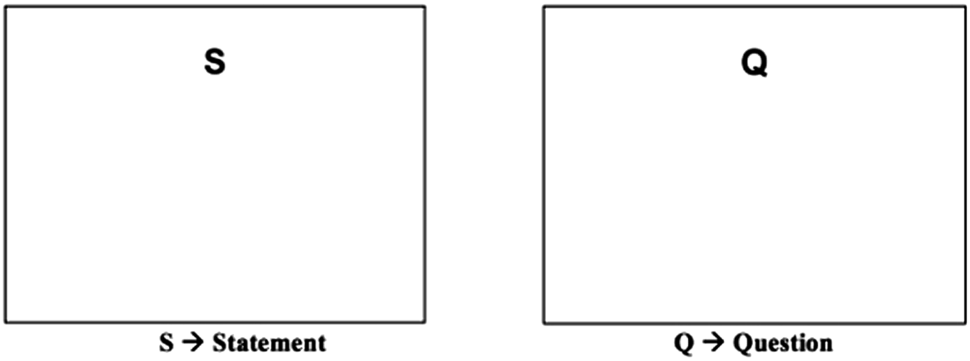

Fig. 1. Sample sentence type screens presented. (a) ‘Statement’ screen: Participants produced declarative statements after seeing an 'S' (b) ‘Question’ screen: Participants produced object wh-questions after seeing a 'Q'.

Methods

Participants

Forty native English speakers participated in exchange for course credit or $10. Two participants were excluded due to poor performance on a recall task (67% correct). Three were excluded due to a failure to understand the task; one was excluded due to a disproportionately high number of outliers in speech onset times (over half of speech onsets were detected as outliers using the Mad-Median rule); four were excluded due to high lexical error rates (three times more than the average error rate). Thus, thirty participants were included in the final analyses.

Materials and design

The experiment was a 2 × 3 within-subjects design with sentence type (declarative versus object *wh*-question) and interfering-word type (subject-related, object-related, or unrelated) as factors. This yielded six conditions, which were rotated across the experiment using a standard Latin Square design. Participants saw a series of two computer screens: (1) An initial ‘sentence type’ screen which told participants whether to produce a statement or a question about the next screen and (2) A ‘critical image’ screen with the characters and actions that participants were instructed to describe. Participants were told that if they had seen an 'S' on the sentence type screen (Fig. 1a), their task was to make a declarative statement about the image screen. However, if they had seen a 'Q' (Fig. 1b), their task was to ask an object *wh*-question about the image screen. As can be seen in Fig. 1, the ‘S’ and ‘Q’ sentence type indicators were centered approximately one inch from the top of the screen. Participants pushed a button on a game controller to advance from the ‘sentence type’ screen to the image screen.

Target image screens consisted of three regions of interest – one around each of the two sets of characters (Fig. 2a–b) and a third around the verb-denoting instrument. As shown in Fig. 2a, the regions of interest around subject and object characters were defined so that there was ample space between the subject/object regions and the region in the middle of the screen, which we defined as the region of interest for the instrument-denoting verb. Characters were left/right balanced such that each character appeared on both the right and left sides of the screen as the subject roughly half the time. In order to ensure that each of the characters chosen for the experiment appeared at least once in each declarative condition (our baseline condition), we added three additional items in the declarative condition to create a total of 33 target items. Crucially, despite this imbalance, character combinations never appeared in the same condition twice so participants were required to produce a novel sentence for each trial. So that participants would not become overly familiarized with the characters, we varied the color of the characters' clothing and also slightly varied their positions within the relevant regions of interest.

Images also included an instrument that signaled which verb participants should produce (e.g. ‘feather’ signals that the verb is ‘to tickle’). In all cases, the grammatical subject was also the agent of the action. The location of the instrument indicated who the subject of the action was. For instance, in (Fig. 2a), the proximity of the feature to the maids indicated that speakers should produce a declarative such as “*The maids tickled the nurses.*” or question such as “*Which nurses did the maids tickle?*”. The instruments were chosen based on prior work showing a strong relationship between each verb and its respective instrument (e.g. Koenig, Mauner, & Bienvenue, 2003; McRae, Hare, Elman, & Ferretti, 2005; Sussman, 2006). Participants were also familiarized with the verbs associated with each instrument beforehand. All verbs were transitive, could have two human arguments, and had a clear semantic distinction between agent and patient. Participants could not anticipate the subject or object character based on the verb. Verbs were repeated 3–4 times during the study, but never in the same condition or with the

Fig. 2a–b. (a) Sample ‘maid’ tickling ‘nurses’ target image screen (b) Sample filler item not containing a verb/instrument image. Participants were free to produce any verb that came to mind.

same sets of characters. Thirty fillers items were also included in the study. On filler trials, participants were also asked to produce either declarative or interrogative sentences (15 each). In some cases, filler items did not include a verb image (Fig. 2b). In these cases, participants were free to produce any verb that came to mind.

To probe which word speakers lexically encoded first after formulating their messages, we also incorporated a picture-word interference (PWI) task (e.g. Schriefers, Meyer, & Levelt, 1990; Roelofs, 1993; Meyer, 1996). At the same time that speakers saw the to-be-described image (0 ms Stimulus Onset Asynchrony), they also heard a phonologically-unrelated interfering word in one of three interference conditions: (a) a word semantically related to the subject in the image, (b) a word semantically related to the object or (c) a word unrelated to both. The interference words and character names were normed in a separate study to ensure that characters and their interfering words were sufficiently related (and unrelated). Following prior work (e.g. Alario, Segui, & Ferrand, 2000; Glaser & Glaser, 1989; La Heij, Dirkx, & Kramer, 1990; Mahon, Costa, Shapiro, & Caramazza, 2002; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007), we defined "relatedness" as a belonging to the same syntactic category (noun) and semantic category (occupations) as the target word. To ensure that our interfering words would cause interference rather than facilitation effects, five additional native American English were asked to perform a continuous association task. Potential target and interference pairs that were too closely associated were eliminated from our materials. Adjectives were used for the unrelated condition.

We compared speech onset times in declarative and interrogative sentences in each of the three interference conditions. To pre-empt the results for the interference manipulation, we found no main effect of sentence type, no main effect of interference type, and no significant interactions. Thus, these results will not be discussed further. We attribute the lack of any significant interference results to the difficulty of our task, relative to other work using picture-word interference tasks. In particular, participants in our study were required to produce two very different syntactic structures on cue. In addition, they were asked not only to pay attention to the interfering word, but also hold it in memory until a subsequent recall task. (In the recall task, participants heard three words and had to indicate if they had heard them in the study. The study included six randomly interspersed recall tasks.). The difficulty of our task was confirmed by converging comments in post-experiment debriefing discussions.

Procedure

Before beginning, participants used a booklet to familiarize themselves with all character and verb images and the associated names/labels. For the production task, participants were instructed to mention all the elements on the screen, to start speaking as quickly as possible, and to use only the names they had learned during the familiarization phase. Moreover, to ensure that participants would specifically produce object *wh*-questions (e.g. rather than subject *wh*-questions, adjunct questions, or yes/no questions), participants were told to ask a question about "who the action is being done to." Because the main experiment was also preceded by ten practice trials, we also had multiple opportunities to correct participants who had formulated a structure other than the desired object *wh*-question.

Visual stimuli were presented on a 22-in. CRT monitor at a resolution of 1024 × 768 pixels (72ppi). Eye-gaze was recorded with an Eyelink II eye-tracker (SR Research, sampling rate 500 Hz). The entire session lasted 45–60 min.

Predictions

In the visual-world eye-tracking paradigm (where speakers describe an image presented on a screen), eye-movements are taken to reflect speakers' visual attention as they produce sentences (e.g. Griffin & Bock, 2000; Bock et al., 2004). Because we are interested in the period in which speakers move from pre-linguistic concepts to linguistic representations, *we focus on eye-movements well before speech onset* (e.g. Griffin, 2001; Gleitman et al., 2007; Konopka & Meyer, 2014).

English declaratives are expected to show the same pattern of eye-movements regardless of whether linguistic encoding is primarily driven by linear word order (positional processing) or syntactic structure (functional processing): Because subjects are the linearly initial element in English declaratives, we should find an initial preference for looks to the subject region, followed by looks to the verb region, and finally looks to the object region.

The crucial condition for teasing apart positional and functional processes in linguistic encoding is the production of English *object wh-questions*. If linguistic encoding is positionally-driven – that is, speakers begin linguistic encoding by assigning a concept to the linearly initial position in the sentence – then, in line with the Obj-Subj-Verb linear word order of object *wh*-questions, we expect speakers to first fixate the object, then the subject, and then the verb. Thus, under a positionally-driven account of linguistic encoding, the order of fixations between declaratives and questions is expected to differ precisely because the linear word order of the structures differs. By contrast, if linguistic encoding is functionally-driven – such that linguistic encoding begins with assigning a concept to the syntactic subject role – speakers should fixate the subject first, regardless of whether they are preparing to produce a declarative or a question. Thus, under a functionally-driven account, the order of fixations prior to speaking in declaratives and questions is not expected to differ.

Results

Data analysis

Although one concern with our design might be the frequency with which participants would actually produce the desired object *wh*-question structure (e.g. relative to subject *wh*-questions), our participants successfully produced the desired object *wh*-question in all but five of the 990 total utterances recorded. All errorful trials (e.g. where participants named characters incorrectly, used the wrong verb, uttered the wrong sentence type) were excluded from both the eye-tracking and speech onset analyses, affecting 4.69% of the data. Disfluent trials (e.g. trials with stutters or filled pauses) were also excluded, affecting an additional 1.05% of the data.[3] Utterance onset times were measured in Praat (Boersma & Weenink, 2009); outliers were determined using the Mad-Median rule (Wilcox & Keselman, 2012) and excluded, accounting for 5.76% of the data. In total, 15.43% of the data was affected; 834 trials were submitted for analysis.

We analyze participants eye-movements in three different ways: (1) To better understand how functional and positional processes interact over the course of linguistic encoding, we analyze the *proportion of looks* to the subject and the object as well as the strength of the subjecthood preference using Subject-Object Advantage Scores (e.g. Arnold et al., 2000; Arnold, Brown-Schmidt, & Trueswell, 2007; Kaiser, 2011). (2) To better understand the timing of functional processes relative to positional ones in linguistic encoding, we analyze the *latency of first looks* to the subject and the object – that is, the time it takes for speakers to first fixate the subject versus linearly-initial object in object *wh*-questions. (3) To investigate whether the patterns in the proportion of looks is linked to speaker-specific differences in *speech onset rates* (i.e. how quickly speakers are to begin speaking), we separate our participants into three groups – fast onsets, medium onsets, and slow onsets – and again analyzed Subject-Object Advantage Scores across speaker groups.

Analyses using the proportions of looks to the subject and to the object were performed separately using the package lme4 (Bates, Maechler, Bolker, & Walker, 2015) in R (version 3.3.2; R Core Team, 2016). In addition, to compare the strength of the preference for the subject *over* the object, we calculated Subject-Object advantage scores by subtracting the proportion of

[3] Additional analyses were performed with disfluent trials included; these showed the same pattern of results.

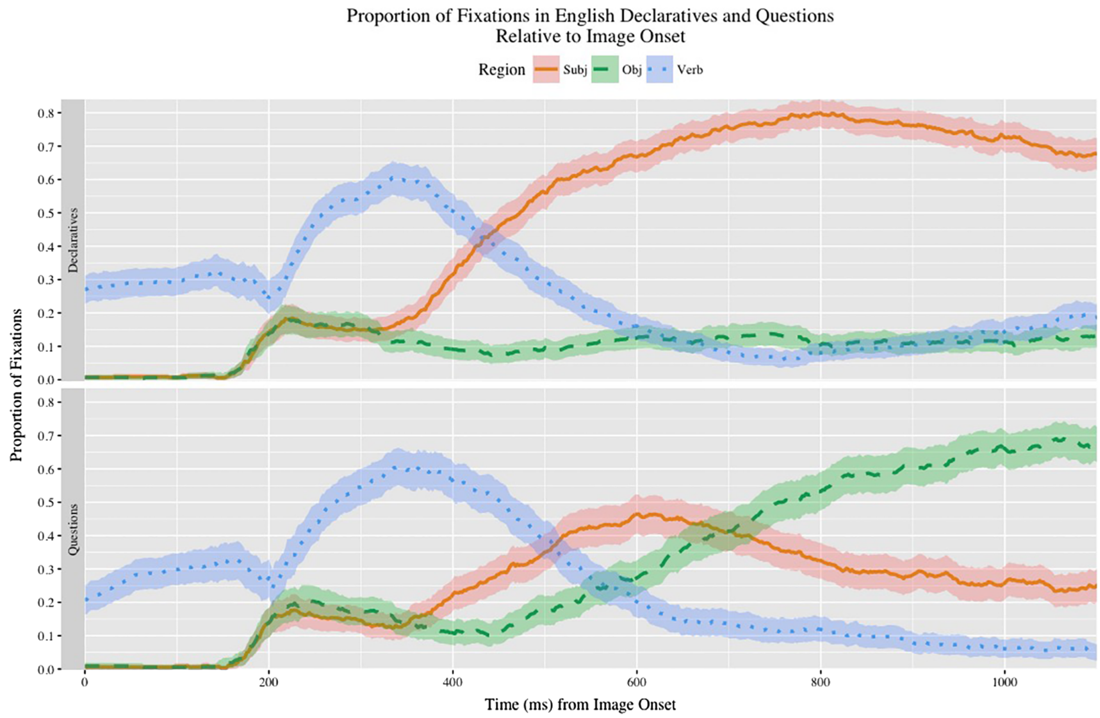

Fig. 3. Looks to subject (solid orange), object (dashed green), and verb (dotted blue) in English relative to critical image onset (i.e. 0 ms indicates onset of critical image). Shaded areas indicate 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

looks to the object from proportion of looks to the subject (e.g. Arnold et al., 2000, 2007; Kaiser, 2011). In this latter analysis, looks to the verb were not considered. Advantage scores were analyzed using Cumulative Linked Mixed Models (CLMM) in the package ordinal (Christensen, 2015).

Following prior work (e.g. Arnold et al., 2000, 2007; Kaiser & Trueswell, 2004; Kaiser, Runner, Sussman, & Tanenhaus, 2009; Kaiser, 2011; Hwang & Kaiser, 2014), eye-movements were binned and separately analyzed in 200 ms time windows, starting from the onset of the to-be-described critical image. The time windows we use for our analyses follow the conventions of prior work (e.g. Griffin & Bock, 2000; Ganushchak et al., 2014; Ganushchak, Konopka, & Chen, 2017). Because we have no *a priori* reason to believe that the process of producing a question (versus a declarative) should shift the relevant windows of analysis, we have chosen to follow prior work in the way that they have de-limited the time windows of analysis for production.

For completeness, both sentence type and interfering word type were always included as fixed effect factors and orthogonally coded in the analyses. But, our eye-movement predictions only focus on differences between declaratives and questions. As previously mentioned, because effects involving interference words have no bearing on our eye-movement predictions, we do not discuss them further. (Recall that we found no significant effects of the interference manipulation in speech onset times.) In other words, we report results that collapse across interfering word types. We started with a fully-crossed random effects model, including by-subject and by-item adjustments to the slope and intercept. Random effects were reduced using model comparison starting with by-item effects; only random effects contributing significantly ($p < .05$) to the model were included (Baayen, 2008). Effects where $|z| > 1.96$ were judged to be significant.

Proportion of looks – across all participants

Fig. 3 plots the proportions of looks to the subject, object, and verb-denoting instrument regions, relative to the onset of the image speakers were asked to describe (i.e. image onset occurs at 0 ms). Note that we are interested in language planning processes that occur well before participants begin speaking; for reference, though, Fig. 4 shows the proportion of looks to the subject, object, and verb, relativized to speech onset in each individual trial. Thus, while Fig. 3 plots proportion of looks entirely before speech onset, Fig. 4 shows the proportion of looks 1000 ms before and 1000 ms after speech onset (see Fig. 5).

Prior work has shown that it takes about 150–200 ms to program and physically launch an eye-movement (Matin, Shao, & Boff, 1993). Thus, what may seem to be a surprisingly high proportion of fixations to the verb region in the 0–200 ms time window are unlikely to be triggered by linguistic processes. This is further supported by the flat/unchanged nature of the looks to the verb region during this time window for both questions *and* declaratives, which mirror the flat proportions of looks to the subject and object. We therefore suggest that the eye-movements during the 0–200 ms time window are best regarded as a theoretically uninteresting side-effect stemming, presumably, from the verb-denoting instrument's location at the center of the screen (Fig. 3) and/or its proximity to the previously presented sentence type indicator.

200–400 ms after image onset: In both declaratives and questions, looks to the verb region start to increase around 200 ms and speakers continue to fixate the verb until around 400 ms. This is in line with prior work in English showing a high proportion of preferential looks to the verb during the earliest points of message formulation for transitive events (e.g. Hwang & Kaiser, 2014; Griffin & Bock, 2000), and is not relevant for the predictions we are investigating in this paper.[4] The proportions of looks to the subject and object increase slightly during this time, but they remain low compared to looks to the verb region. There is no main effect of sentence type during this time slice either in

[4] While the underlying reason for these early looks to the verb remain an open question, this pattern should not be taken as unpredicted or surprising given what has been reported by prior work. For instance, other work showing this same pattern of results has suggested early verb fixations to be related to the centrality of the verb in determining the message-level representation of the sentence (Hwang & Kaiser, 2014). The design of our images raises an additional possibility; namely, that participants were forced to fixate the verb first because they were explicitly told that characters closer to the verb instrument were subjects of the action. The results reported here are consistent with both possibilities. Thus, clarifying the role of the verb in production more generally remains an avenue of future investigation, but is beyond the scope of this paper.

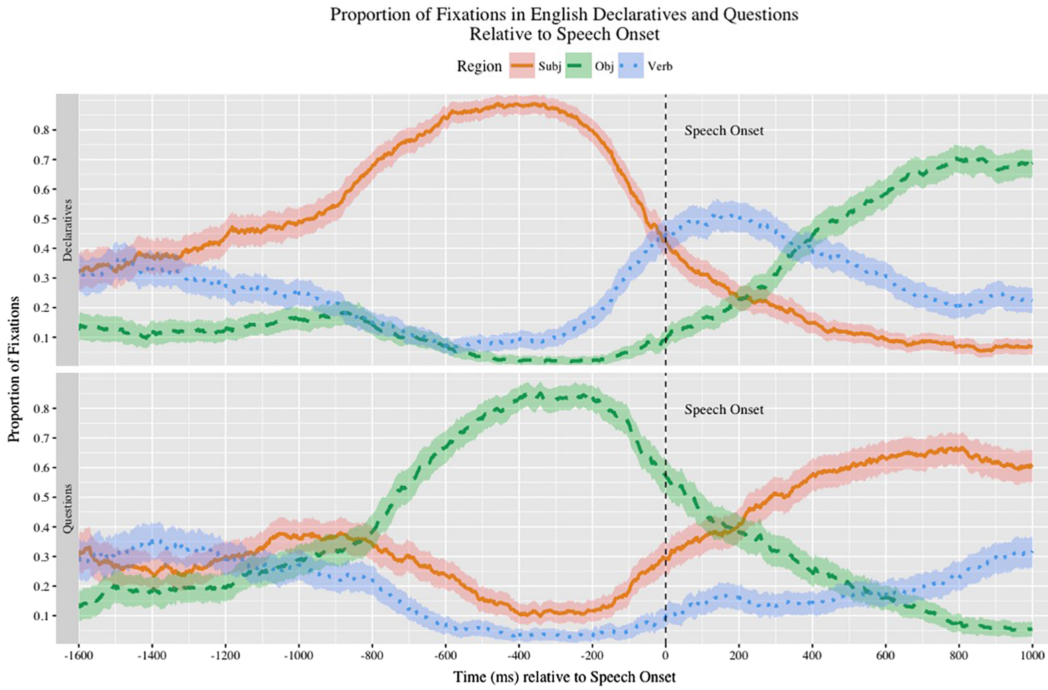

Fig. 4. Looks to subject (solid orange), object (dashed green), and verb (dotted blue) in English relative to speech onset (i.e. 0 ms indicates speech onset) of each individual trial. Shaded areas indicate 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

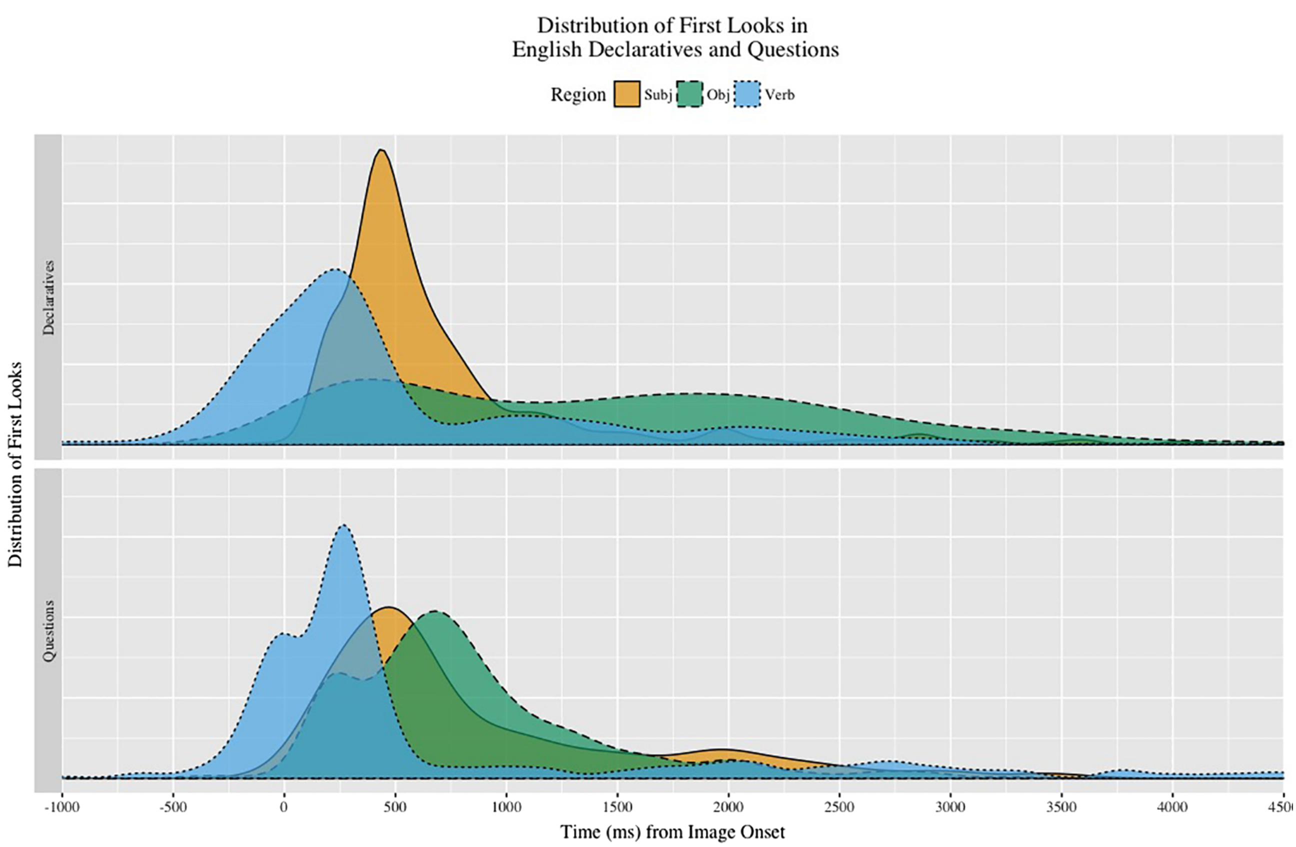

Fig. 5. Shows density plot of the latencies of first looks to subject (solid/orange), object (dashed/green), and verb (dotted/blue) in English. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Mean, standard error, and median latencies in milliseconds for the first looks into the subject, object, and verb interest areas.

Interest area	Sentence type	Mean latency (ms)	SE latency (ms)	Median latency (ms)
Subj	Declarative	718.980	30.247	497.0
	wh-Question	887.548	37.464	575.5
Obj	Declarative	1486.004	49.085	1440.0
	wh-Question	830.359	28.076	721.0
Verb	Declarative	552.875	41.728	252.0
	wh-Question	679.354	55.499	256.0

looks to the subject ($\beta = -.004$, SE = .65, $|z| = .006$) or object ($\beta = 1.14$, SE = .73, $|z| = 1.56$). Using the Subject-Object advantage scores, we also detect no difference in the strength of the subject-object preference across sentence types: ($\beta = -0.21$, SE = .63, $|z| = 0.33$).

400–600 ms after image onset: Differences between subject and object looks in declaratives vs. object *wh*-questions begin to emerge during this time window. In declarative sentences, the proportion of looks to the subject increases rapidly while the proportion of looks to the object remain low. Recall that looks to the object vs. the subject in *wh*-questions (as compared to declaratives) is the crucial testing ground that will allow us to pull apart positional and functional approaches to linguistic encoding. Indeed, what we see is that in *wh*-questions – where the object linearly precedes the subject – looks to *both* the object and the subject begin to increase during this time window. Critically, in object *wh*-questions the proportion of looks to the *subject* appears higher than the proportion of looks to the object, even though the subject is *not* the linearly initial element in the sentence.

Statistical analyses of looks to the subject and object in this time window correspond to what is shown in Fig. 3, though differences in this time window only reach marginal significance. In particular, we find more looks to the subject in declaratives ($\beta = .91$, SE = .69, $|z| = 1.33$) and a corresponding marginal effect of sentence type ($\beta = .81$, SE = .60, $|z| = 1.36$) in the Subject-Object advantage score. In the analyses of looks to the object, we find more looks to the object in object *wh*-questions ($\beta = -.83$, SE = .53, $|z| = 1.56$); again, these differences are marginally significant during this time window. Thus, differences between declaratives and questions begin to emerge during the 400–600 ms time window after image onset, but do not yet reach significance.

600–800 ms after image onset: Patterns that began to emerge in the 400–600 ms time window reach significance by 600–800 ms after image onset. Specifically, in declarative sentences, looks to the subject region continue to rise steadily, reaching a peak at roughly 800 ms, while looks to the object remain low. In contrast, *wh*-questions show a very different pattern: By 600 ms, looks to the subject reach a peak and begin to decline; at the same time, looks to the object steadily increase suggesting that speakers are shifting their attention from the subject to the object.

Statistical analyses confirm what is shown in Fig. 3: We find a main effect of sentence type such that there are significantly more looks to the subject in declaratives than in questions ($\beta = 1.42$, SE = .71, $|z| = 2.02$) and correspondingly, significantly more looks to the object in questions than in declaratives ($\beta = -1.71$, SE = .61, $|z| = 2.78$). These differences are also reflected in the advantage scores, where we find a significant main effect of sentence type ($\beta = 1.72$, SE = .64, $|z| = 2.67$), meaning that the subject preference in declaratives is larger than the subject preference in questions.

800–1000 ms after image onset: Differences between declaratives and questions extend into the 800–1000 ms time window. After 800 ms, looks to the subject start to decline in declarative sentences, but in questions, looks to the object continue to rise reaching a peak at roughly 1000 ms. There is a main effect of sentence type (Looks to Subj: $\beta = 1.46$, SE = .5, $|z| = 2.93$; Looks to Obj: $\beta = -2.41$, SE = .58, $|z| = 4.14$), confirming these differences. Again, Subject-Object advantage scores show a main effect of sentence type such that the subject preference in declaratives is larger than in questions ($\beta = 1.68$, SE = .55, $|z| = 3.08$).

Latency of first looks

Analyses over the proportion of looks to the subject versus the object suggest an interesting interaction between functional and positional processes during linguistic encoding of object *wh*-questions. To better understand the nature and timing of this interaction, we directly compare the time it takes for speakers to *first* fixate the subject versus object (e.g. Kaiser & Trueswell, 2004; Brown-Schmidt & Tanenhaus, 2006). We therefore compute the mean, standard error of the mean, and median onset times of first looks to the subject versus object. These are reported in Table 1.

Note that although the general pattern of mean latencies conforms to what we expect given the data showing the proportion of looks (Fig. 3), the mean latencies appear to be larger/slower than what is reported in Fig. 3. As can be seen in Fig. 6, these extended latencies are likely driven in large part by outliers in the data. A look at the *median* onset times reported in Table 1 reveal latencies that show the same pattern of results as mean times, but appear to be much more in line with what is expected given Fig. 3. We, therefore, used a percentile bootstrap (`dmedpb`, Wilcox, 2017; Wilcox & Rousselet, 2017) to separately compare the median latency of looks to the subject and object across sentence types. We believe that this comparison is preferable because medians are more robust to deviations from normality that are driven by outliers; this is precisely the case of the data here.[5]

Comparing across sentence types, we find no difference in the latency of looks to the subject across sentences types (psi-hat = −92.114; CI = (−126.122, 11.410); p = 0.07), suggesting that speakers first fixate the subject at roughly the same time in declaratives and object *wh*-questions. Thus, regardless of whether the subject is the sentence-initial element or not, first fixations to the subject occur early in both sentence types. As expected based on the eye-movement data, we find a significant difference in the latency of first looks to the object: Speakers looked to the object significantly sooner when preparing to produce a question than when preparing to produce a declarative (psi-hat = −73.778; CI = (−207.529, −13.067); p = 0.012).

In sum, data from the median latencies of first looks to the subject and object converges with what emerged from analyses over the proportion of looks to the subject and object. As expected, the linear word order of the utterance did affect the time course of object planning – speakers attended to the object sooner when it occurred earlier (e.g. in object *wh*-questions) than when it appeared later (e.g. declaratives) in the sentence. But, looks to the subject of the sentence emerge during the same time window – regardless of whether the subject is the linearly initial element in the sentence (e.g. in declaratives) or not (e.g. in object *wh*-questions).

Thus, subjects are privileged regardless of linear order (*wh*-question or declarative), whereas the fate of objects depends on the linear order of the utterance.

Proportion of looks – by speech onset groups

It is well-known that the scope of planning (i.e. the size of the incrementally planned ‘chunks’) can vary depending on language-specific

[5] For comparison, additional analyses were performed over the means using linear mixed effects regressions (Bates et al., 2015). These yielded the same pattern of results: The latency of first looks to the subject did not differ between sentence types ($\beta = 133.18$; SE = 44.21; $|t| = 1.28$; $p = 0.207$); but looks to the object were significantly later in declaratives than in object *wh*-question ($\beta = -620.59$; SE = 40.54; $|t| = 5.828$; $p < 0.001$).

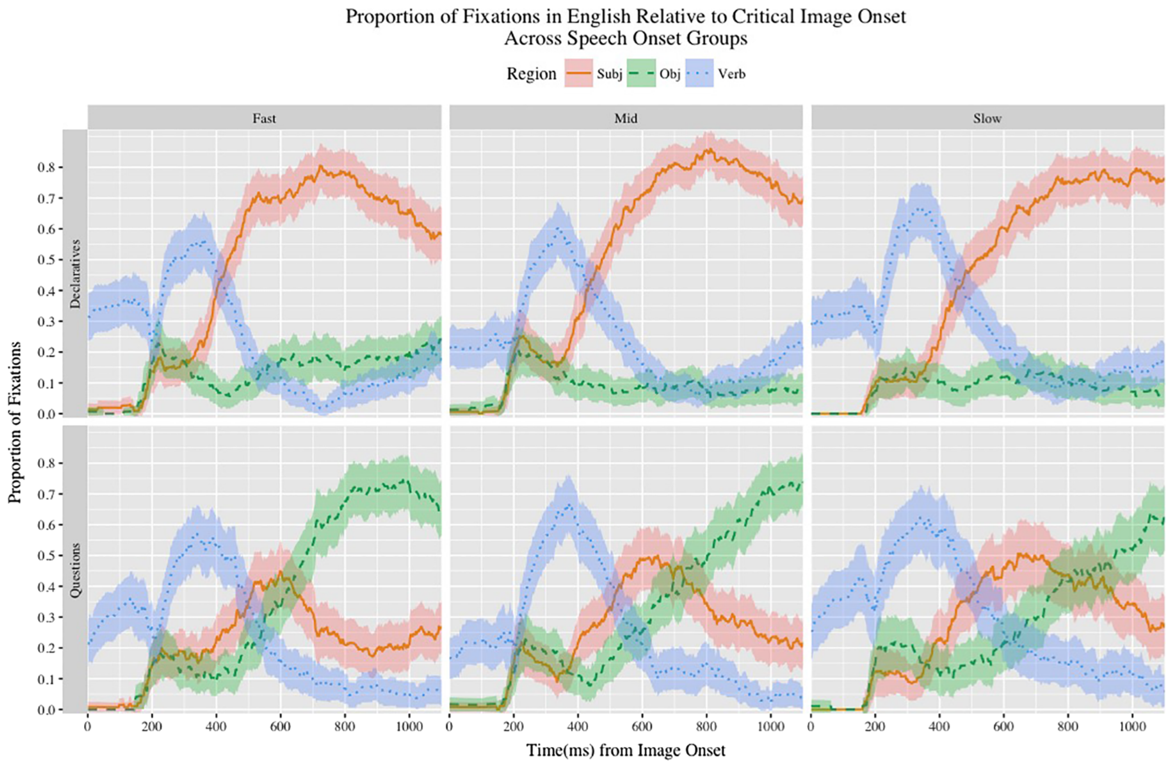

Fig. 6. Proportion of looks to subject (solid orange), object (dashed green), and verb (dotted blue) in English, broken down by speech rate. 0 ms represents onset of critical image. Shaded areas indicate 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

factors (e.g. Griffin, 2003; Brown-Schmidt & Konopka, 2008) as well as on factors like cognitive load (e.g. Ferreira & Swets, 2002; Wagner, Jescheniak, & Schriefers, 2010) and message complexity (e.g. Kuchinsky, 2009; Bock et al., 2004). More specifically, speakers flexibly adjust the scope of planning in response to increased cognitive load and/or message complexity. Because the size of a planning unit has implications for the starting point of linguistic encoding, a possible concern is that what appears as evidence for competition between the subject and linearly first object at the level of linguistic encoding, might alternatively be attributable to speaker-specific differences in the scope of planning. For instance, under increased cognitive load, some speakers might narrow the scope of their planning (e.g. Ferreira & Swets, 2002; Wagner et al., 2010); in this case, they may take a highly incremental approach to planning such that they plan only the linearly initial element of the sentence (e.g. Gleitman et al., 2007; Brown-Schmidt & Konopka, 2008; Brown-Schmidt & Tanenhaus, 2006). Here, speakers who are faster to begin uttering their sentences might disproportionately drive the pattern of fixations in object *wh*-questions. It is also possible that under increased cognitive load, some speakers may develop a more fully 'fleshed-out' representation of their utterance. In this latter case, speakers might widen their scope of planning and plan both the subject and the object of the utterance. In other words, speakers who are slower to begin uttering their sentences might be disproportionately responsible for the pattern of fixations in object *wh*-questions. Regardless of these possibilities, both options link differences in eye-movements between declaratives and object *wh*-questions not to 'competition' between functional versus positional encoding processes, but to individual differences in the rate at which speakers begin uttering their sentences.[6]

We note, though, that even if one assumes that the looks to the subject and looks to the object were driven by different groups of speakers (i.e. that our participants are bimodally distributed such that some speakers consistently look to the subject first while others consistently look to the object first) our results – namely, that at least some speakers fixate the subject first – are still striking. In particular, even if our results are driven by bimodality in the data, this does not negate that a meaningful proportion of our speakers do not linguistically encode following the linear word order of the utterance – that for these speakers, structural planning is the 'default' even when the deck is stacked against the structural approach. Rather, that bimodality in our data would suggest is simply that future work should systematically investigate why speakers might take one approach over the other.

Nonetheless, to investigate the extent to which individual differences in speech onset rates might drive eye-movement patterns, we performed a follow-up analysis akin to the one reported by Wagner et al. (2010). Their study directly investigated the conditions which might contribute to changes in scope of planning and their dependent variable was speech onset times. But, they were similarly concerned that their results could be modulated by the individual speech onset rates of their participants. Wagner et al. (2010) divided their speakers into two groups – a fast and slow group – using a median split. We choose to divide our speakers into three groups – fast ($\bar{x}$ = 941.30 ms; SE = 20.76), medium ($\bar{x}$ = 1279.00 ms; SE = 23.64), and slow ($\bar{x}$ = 1559.38 ms; SE = 29.02) groups – based on their average speech onset times. This was done because a split into thirds would more effectively capture speaking rate differences given the right-skewed distribution of our speech onset data (Gelman & Park, 2008).[7] In addition, because analyses of speech onset times detected no main or interaction

[6] We thank Benjamin Swets for this suggestion.

[7] We also conducted additional analyses based on a median split of speech onset times, as done in Wagner et al. (2010). Results via median split showed the same pattern: We detected no interaction effects involving speech rate.

effects involving either sentence or distractor type, averages for speech onset times were computed by collapsing across both of these factors.

Statistical analyses over the Subject-Object advantage scores were performed as before. However, speech rate group was included as a fixed effect and as a random by-item effect (it was not included as a random effect by-subjects because speech onset group is a between-subjects factor).

We found no significant interactions involving speech rate before 600 ms (all $|z| < 0.949$; all $p > 0.3424$). However, we do find significant speech rate interactions during 600–800 ms ($\beta = 2.14305$; $SE = 1.01238$; $|z| = 2.117$, $p = 0.0343$) and 800–1000 ms ($\beta = 1.2098$; $SE = 0.5511$; $|z| = 2.195$, $p = 0.0281$) time windows. In both, we find that slower speakers show a significantly stronger subject preference than fast or medium speakers combined. Thus, although fast and medium speakers also prefer the subject over the object, it appears that slower speakers show a larger subject-preference. This result is in line with the possibility that slower speakers may adopt a wider scope of planning to develop a more 'fleshed-out' representation before speaking.

Crucially, though, speakers in all groups show the same general pattern of eye-movements with each other (Fig. 6) and by extension, with the trend originally observed in Fig. 3. In declaratives, speakers in all groups looked primarily to the subject, while looks to the object remain comparatively low. In questions, we also find that speakers in all groups behave similarly. In the 400–600 ms window, they look first to the subject while looks to the object rise simultaneously during that same time window.

In sum, we find the same pattern of eye-movements regardless of whether speakers are quick to begin speaking or whether they are slow to begin speaking. Specifically, it is ***not*** the case that participants who begin speaking quickly drive looks to the object while participants who begin speaking more slowly drive looks to the subject or vice versa. Rather, the same general pattern of eye-movements holds across participants regardless of how quickly they begin speaking.

Experiment 1 discussion

To move from thoughts to utterance, speakers must translate their thoughts into speech. This requires that each concept be assigned (i) a grammatical role – through functional processing, and (ii) a linear order in the sentence – through positional processing. Despite prior cross-linguistic work (Griffin & Bock, 2000; Gleitman et al., 2007; Myachykov et al., 2011; Norcliffe et al., 2015), an open question is whether linguistic encoding starts with positional processing, where concepts are encoded to linear word order positions in the sentence, or whether encoding begins with function assignment, where concepts are assigned a 'slot' in the syntactic structure of the utterance. The current work investigates this question using object *wh*-questions (e.g. '*Which nurses did the maids tickle?*'). Crucially, in these sentences, the object linearly precedes the subject, which allows us to test whether grammatical functional assignment can drive the course of linguistic encoding even in spite of competing constraints stemming from the linear word order of the utterance. In particular, we are interested in whether speakers will (i) plan the syntactic subject first (as functional processing would predict) even when it is not the element that gets uttered first or whether they will (ii) plan the linearly-initial element, the one that they need to produce first (as positional processing would predict).

During linguistic encoding of declarative sentences, eye-movements show that the speakers in our experiment treat the subject as the starting point for encoding and do not consider the object as a potential starting point. In particular, the proportion of looks to the object remains low while the proportion of looks to the subject rose rapidly. This is expected given that there is no tension between functionally- and positionally-driven approaches to linguistic encoding: The subject is also the linearly first element in a declarative sentence.

In our critical condition of interest, though – object *wh*-questions – we found that even when speakers knew that they would have to *say* the object of the sentence before saying the subject, they *linguistically encoded* the subject of the sentence before encoding the object. This was evidenced by differences in the *strength* of the subject preference between declaratives and questions – a pattern that held for all participants, regardless of their individual speaking rates – and, was corroborated by analyses of the timing of the first looks to the subject versus object in object *wh*-questions. These results suggest that subjecthood is privileged over the linearly first element of an utterance and by extension, that speakers may initially 'default' towards functional processes to guide the starting point of linguistic encoding.

In addition, our results show that speakers fixate the object relatively soon after fixating the subject in object *wh*-questions – presumably to encode the linearly-initial object. That speakers turn their attention to the linearly-initial object relatively soon after attending to the subject provides further evidence that functional and positional processes are, distinct, but tightly-coordinated aspects of linguistic encoding. In particular, this pattern of results shows what the time course of non-serial planning – wherein speakers do not have to wait to fully encode the subject to begin encoding the linearly initial object – can look like in real time.

In sum, the results of Experiment 1 suggest that functional and positional processes can play independent, though tightly-linked roles in linguistic encoding. Crucially, though, functional processes play a privileged role in linguistic encoding. Our results indicate that the first step in the transition from abstract conceptual representations to highly structured linguistic ones starts with encoding of the *subject*, even when speakers have to actually articulate the *object* first (i.e. when the subject is not the linearly initial element in the sentence).[8]

Experiment 2: Investigating focus effects using Mandarin Chinese

While Experiment 1 shed light on the timing and interaction between functional versus positional processes during linguistic encoding, a potential concern has to do with the illocutionary, semantic and pragmatic differences between declaratives and questions. *Wh*-phrases are informationally focused elements, and signal that a speaker is seeking information they do not yet know. As previously noted, one potential complication associated with investigating the production of sentences with non-canonical word orders (Myachykov et al., 2011; Norcliffe et al., 2015; *inter alia*) is the role of discourse-pragmatic factors. In fact, the role of discourse/pragmatic effects in the process of production planning is not yet well-understood. In one of the few eye-tracking production studies related to questions, Ganushchak et al. (2014) asked participants a neutral-focused (*'What is happening here?'*), subject-focused (*'Who is stopping the truck?'*), or object-focused (*'What is the policeman stopping?'*) *wh*-question. After hearing the question, participants saw a scene that they described in response to the question. Because participants were instructed to respond using complete sentences, the structure of the response was always the same regardless of the type of question participants heard: *'The policeman is stopping the truck'*.

That work found that information focus could affect eye-movements during the window typically associated with linguistic encoding (e.g. 400–1000 ms after image onset). Critically, though, because participants in Ganushchak et al. (2014) only produced *declarative responses to questions* (i.e. questions were used strictly to elicit participants' declarative responses), those results do not speak directly to

[8] Our results – namely, that speakers encode the subject before the object even in an object-initial *wh*-question – also appears to be in line with what might be predicted by Chomskyan transformation-based grammars (e.g. Chomsky, 1977; Ross, 1967). But, locating the source of the subjecthood preference is beyond the scope of the current paper.

role of informational focus during *question* production, itself. Specifically, it is not clear whether those results are due to planning processes intrinsic to production or whether these were the by-product of earlier comprehension processes, such as identifying the focused element.

Relevant for us, Ganushchak et al.'s (2014) work opens up the possibility that the results observed in Experiment 1 might not be exhaustively attributed to the interaction between positional and functional constraints on linguistic encoding, but also to informational focus effects associated with question formation. The main aims of Experiment 2 are twofold: First, we aim to investigate the possibility that the results observed in Experiment 1 might not have been exhaustively attributed to the interaction between positional and functional constraints on linguistic encoding, but also to informationally focused *wh*-phrases in *wh*-questions. Second, current understandings of the role of information focus during real-time language production is rather limited: Very little work has investigated the role of information focus on the process of production planning (though see Ganushchak et al., 2014). Given the state of the literature in this area, we also hope to clarify the extent to which information focus effects (operationalized in this work as informationally focused *wh*-phrases) can affect the process of linguistic encoding, itself.

In Experiment 2, we compare the production of declaratives and object *wh*-questions in Mandarin Chinese, a language where declaratives and questions share the same linear word order (2a-b).

(2a) Mandarin Declarative:	厨师们	枪毙了	护士。
	The chefs	shot	the nurses.
(2b) Mandarin *Wh*-Question:	厨师们	枪毙了	哪个护士?
	The chefs	shot	which nurses?

Because Mandarin declaratives and object *wh*-question differ only in terms of their information focus properties (i.e. subjecthood and linear word order converge in both declaratives and object *wh*-questions in Mandarin), we can hold linear word order and syntactic role assignment constant.[9] Thus, functional and positional processes should unfold in the same way across sentence types; any differences in speakers' eye-movements should not be attributable to the interaction of the factors guiding *these* processes, but to something else. In the case of object *wh*-questions, information focus appears to be a likely candidate (Ganushchak et al., 2014).

Methods

Participants

Forty-eight native speakers of Mandarin Chinese, born and raised in Mainland China, participated. Of these 48, one speaker was excluded from analysis due to sleepiness. An additional nine were excluded due to a disproportionately high number of naming errors. Finally, three were excluded due to poor performance on the auditory recall tasks (less than 66% accuracy). A total of 35 participants were included in the final analyses.

Materials and design

The design of Experiment 2 was the same as in Experiment 1. However, individual items (e.g. verb instruments, characters, etc.) were modified to account for lexical differences between Mandarin and English. As in Experiment 1, participants also heard subject-related, object-related, and unrelated interfering words at the same time that they saw the critical images. Interfering words in Mandarin were selected in a separate norming experiment.

To again pre-empt our results, we find no main effects of sentence type or interfering word type in speech onset latencies in Mandarin. There were also no significant interactions. Thus, we do not discuss the picture-word interference paradigm or the (lack of) results further.

Procedure

The procedure was the same in Experiment 1, except that rather than seeing an 'S' or 'Q' on the screen preceding the to-be-described image, participants saw '陈述' ('statement') or '设问' ('question') to indicate whether they should produce a statement or a question, respectively.

Results

Data analysis

Out of 1155 total sentences, trials containing errors and/or disfluencies were excluded from analysis, affecting 16.62% and 5.97% of the data, respectively. As in English, outliers in speech onset times were determined using the MAD-median rule and excluded from analysis; this affected 10.37% of the remaining data. Statistical analyses over proportion of looks to the subject, the object, and Subject-Object Advantage scores were done as before. We again performed two analyses related to the proportion of looks – the first, *across all participants*, and the second, with speakers divided into a fast, medium, and slow group by *speech onset rates*. Given the eye-movement patterns observed for Mandarin speakers, analyses over the latency of first looks to the subject versus object (see Experiment 1) did not seem necessary.

Proportion of looks – across all participants

Fig. 7 shows eye-movements for declaratives and object *wh*-questions for Mandarin speakers relative to image onset. We are not interested in eye-movements after speech onset, but for completeness, these are shown in Fig. 8, where 0 ms is relativized to the speech onset time of each individual trial.

As in Experiment 1 (English), **200–400 ms after image onset**, speakers fixate the verb in declaratives and questions. **400–600 ms after image onset**, speakers rapidly shift their gaze towards the subject of the sentence in declaratives and questions alike; we find no difference in the proportion of looks to the subject between sentence types during this time window ($\beta = .91$, $SE = .69$, $|z| = 1.33$). Declaratives and object *wh*-questions in Mandarin continue to pattern alike through 1000 ms after image onset. In both sentence types, speakers preferentially fixate the subject, while looks to the object remain comparatively low. This is reflected statistically in analyses over the proportion of looks to the subject and looks to the object, where we detect no significant effect of sentence type (all $|z| < 1.01$) in any time window after speech onset. Subject-Object advantage scores (all $|z| < 1.49$) also suggest there was no difference in the strength of the subject preference between declaratives and object *wh*-questions.

[9] In English object *wh*-questions, the wh-phrase '*Which nurse*', is obligatorily pronounced at the start of the sentence, but is interpreted as the object of the verb 'tickle'. Therefore, English *wh*-questions are said to involve a dependency between (i) where the wh-phrase is pronounced and (ii) where it is interpreted. Since Mandarin *wh*-phrases are pronounced and interpreted in the same position, Mandarin questions do not appear to involve any *wh*-dependency. However, a large body of theoretical work (e.g., Aoun & Li, 1993; Cheng, 1991; Huang, 1982) and recent experimental evidence (e.g. Aoshima et al., 2004; Ueno & Kluender, 2009; Xiang, Dillon, Wagers, Liu, & Guo, 2014; Xiang, Wang, & Cui, 2015) suggests that even when *wh*-questions do not appear to involve a dependency, *wh- in situ* languages like Mandarin actually involve a covert dependency analogous to overt ones found in languages like English. As will be shown below, the question of whether Mandarin object *wh*-questions do or do not contain a covert dependency does not bear on our interpretation of the current work.

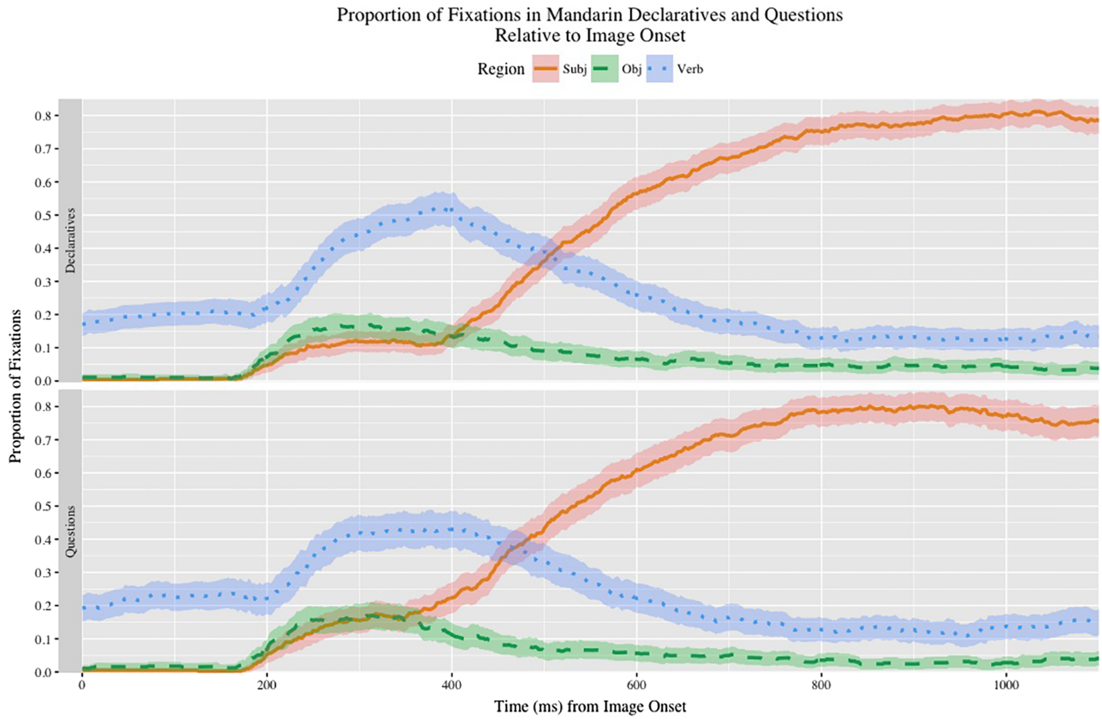

Fig. 7. Shows looks to subject (solid orange), object (dashed green), and verb (dotted blue) in Mandarin. 0 ms represents onset of to-be-described image. Shaded areas indicate 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

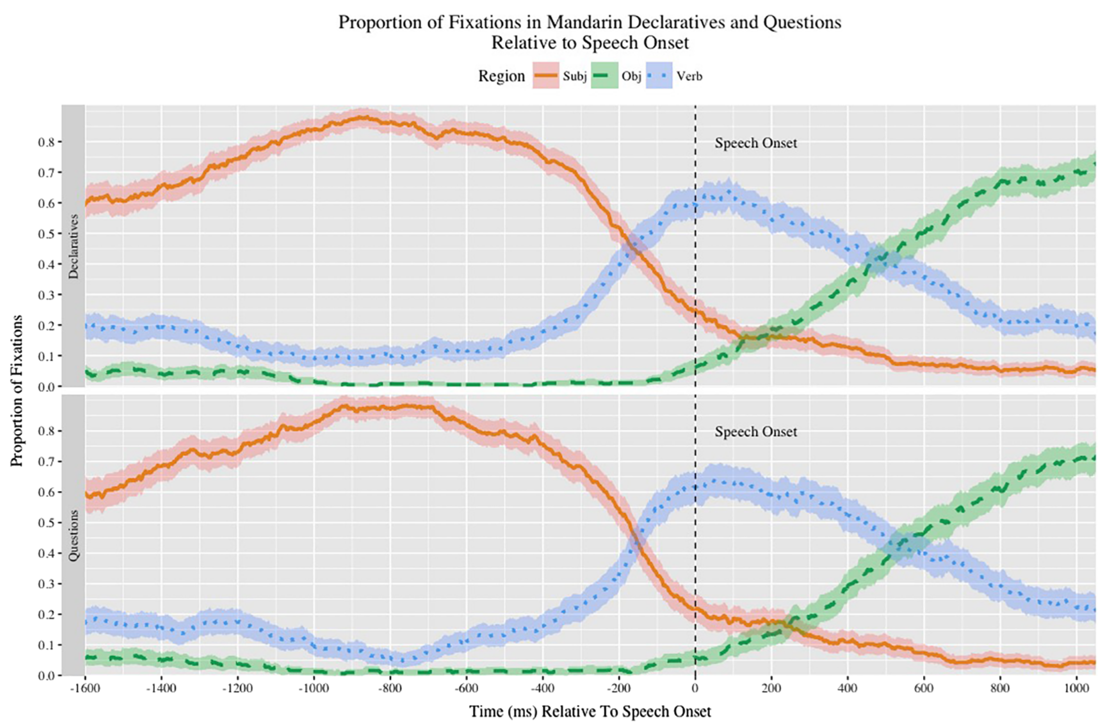

Fig. 8. Shows looks to subject (solid orange), object (dashed green), and verb (dotted blue) in Mandarin. 0 ms represents speech onset in each individual trial. Shaded areas indicate 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Proportion of looks – by speech onset groups

Given the pattern of eye-movements observed here, we believed that it was unlikely for our results to be modulated by differences in speech rates across speakers. However, we performed speech rate analyses akin to the one reported in Experiment 1. We, again, divide our Mandarin speakers into fast ($\bar{x}$ = 1213.77 ms; SE = 30.42), medium ($\bar{x}$ = 1631.26 ms; SE = 36.48), and slow ($\bar{x}$ = 2212.06 ms; SE = 44.59) groups based on averaged speech onset times collapsed across sentence and interference type. As shown in Fig. 9, the same general pattern of eye-movements that emerged when we analyzed across all participants in Experiment 2 holds for fast, medium, and slow speakers. No significant interactions involving speech onset rates were detected during any time window in Mandarin (all $|z| < 1.68$). This suggests that our results in Experiment 2, like in Experiment 1, were not

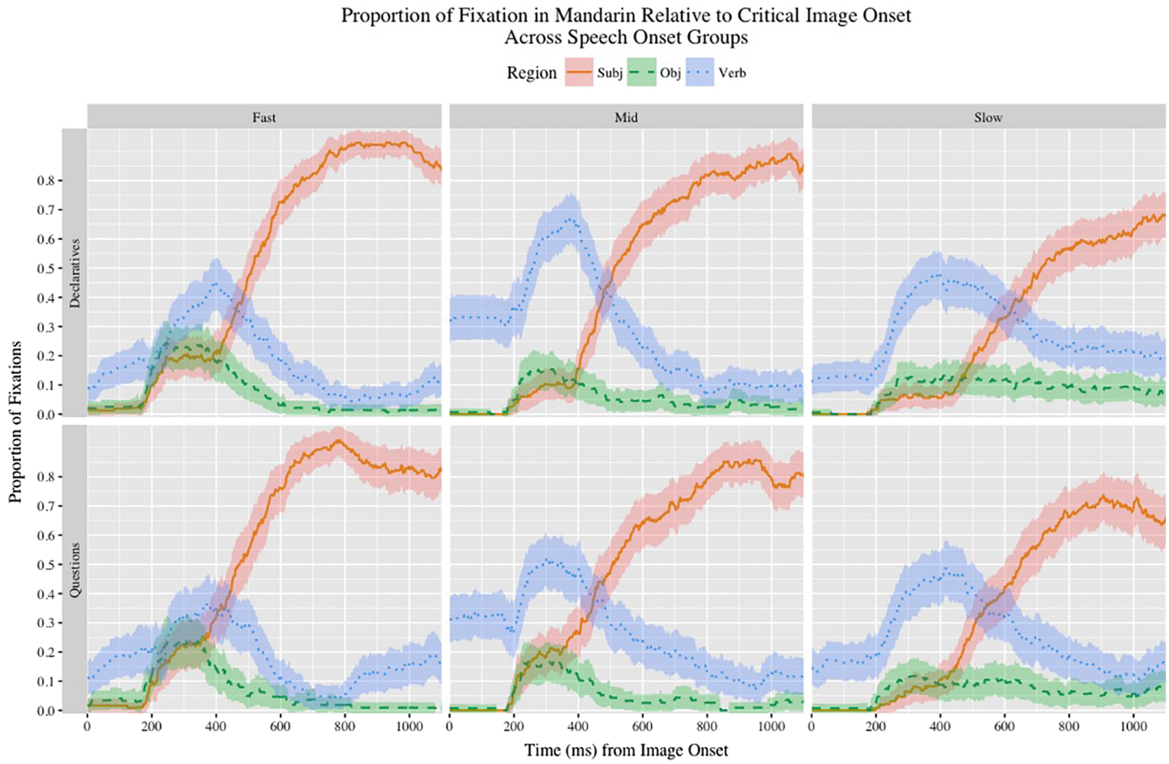

Fig. 9. Shows proportion of looks to subject (solid orange), object (dashed green), and verb (dotted blue) in Mandarin, broken down by speech onset group. 0 ms represents onset of critical image. Shaded areas indicate 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

disproportionately driven by a subset of participants.

Experiment 2 discussion

There were two main motivations for Experiment 2: First, one concern with the results of Experiment 1 was that what we interpreted as competition between looks to the subject versus the object in English object *wh*-questions was driven not by effects of subjecthood versus linear word order, but by effects of subjecthood versus information focus. Thus, we wanted to investigate the extent to which the results of Experiment 1 might have been confounded by effects of information focus. Second, existing work on the role of information-structural factors on language production is scarce. As far as we know, only Ganushchak et al. (2014) investigated effects of information focus on production, and that was in the context of *responding* to *wh*-questions with declaratives. Thus, we wanted to investigate the extent to which the presence of an informationally focused element can influence eye-movements during linguistic encoding.

Experiment 2 does not find any evidence that focus effects stemming from illocutionary differences between declaratives and questions drive eye-movements during linguistic encoding. This means that the results of Experiment 2 largely exclude the possibility that information focus affects eye-movements during linguistic encoding and thus argue against the possibility that the results observed in Experiment 1 (*wh*-question production in English) were confounded with information focus. In other words, Experiment 2 provides evidence that the results from Experiment 1 were driven primarily by the interaction between subjecthood and linear word order, not by information focus.

Notably, the results reported here differ in some respects from what is reported by Ganushchak et al. (2014), which also used an eye-tracking-during-production visual world paradigm to investigate the role of focus on production planning. Unlike participants in our work (who produced questions), participants in Ganushchak et al. (2014) were *asked* informationally focused questions (e.g. '*Who is the policeman stopping?*') but *produced* only declarative sentences (e.g. '*The policeman is stopping the truck.*') responding to those questions. Their results showed that during linguistic encoding (i.e. 400 ms after critical image onset) participants began preferentially fixating the subject character in the subject-focused condition and began preferentially fixating the object character in object-focused conditions. In other words, their results suggest that information focus can impact eye-movements during linguistic encoding. Unlike Ganushchak et al. (2014), we do not detect any effect of information focus on eye-movements during this same time window. We believe that one explanation for the disjunction between our work and Ganushchak et al.'s (2014) may be the way in which these focus effects were experimentally induced. Consequently, the different eye-movement patterns observed by Ganushchak et al. (2014) across their focus-conditions may not be intrinsic to the production of questions themselves. Rather, eye-movements in their study reflected changes in the discourse context by way of comprehending (or recalling) the question that participants were asked to answer.

In sum, by separating out effects of linear word order and information focus on linguistic encoding, the results of Experiment 2 largely exclude the possibility that information focus affects eye-movements during linguistic encoding. Instead, the results of Experiment 2 suggest that the results observed in Experiment 1 were driven primarily by the interaction between functional and positional processes in linguistic encoding; they were not confounded by information focus. Furthermore, taken in the context of prior work by Ganushchak et al. (2014), our results also suggest that the question of how illocutionary differences (e.g. between declaratives and questions) can directly affect production planning is still an open one that requires further investigation.

General discussion

In order to go from an abstract message to a linguistic utterance, speakers have to integrate unordered concepts from the conceptual level of representation into a highly-structured syntactic frame at the linguistic level of representation. This process of translating messages into linguistic utterances is known as linguistic encoding. It requires each abstract concept of a message to be assigned both (i) a slot in the linear order of the sentence *(positional processing)*, and (ii) a role in the hierarchical syntactic representation of the sentence *(functional processing)*. But, the how the process of linguistic encoding unfolds – and in particular, to what extent the starting point of linguistic encoding is driven by positional versus functional processes – is not yet well-understood. This is the outstanding question that the present work seeks to address. Under positionally-driven accounts, linguistic encoding begins with direct assignment of a salient concept to the linearly-initial position in the utterance, with grammatical function assignment following from the initially-selected concept (e.g. Gleitman et al., 2007; Brown-Schmidt & Konopka, 2008; among others). Under functionally driven accounts, linguistic encoding begins with assigning a concept to the function of syntactic subject, with the linear position of the syntactic subject being determined afterwards (e.g. Griffin & Bock, 2000; Lee et al., 2013; Momma et al., 2016).

To tease apart these two possibilities, we use the visual-world eye-tracking paradigm to investigate the real-time production of object *wh*-questions in English (e.g. '*Which chefs did the nurses tickle?*'; Experiment 1) and Mandarin Chinese (Experiment 2). In English, object *wh*-questions are a construction that effectively 'stacks the deck' against a functionally-driven approach to linguistic encoding. Specifically, in contrast with declaratives (e.g. *'The nurses tickled the chefs.'*), a critical feature of English object *wh*-questions is that the subject is not the linearly-initial element of the sentence. This means that this construction provides an ideal testing ground for separating out (i) positional processes, which predict that speakers should first encode the conceptually salient linearly-initial element (i.e. the object) from (ii) functional processes, which predict that speakers should first encode the syntactically privileged subject. On the one hand, given that speakers have to utter the object of a *wh*-question first, an 'obvious' approach to linguistic encoding appears to be the maximally efficient, positionally-driven one: Speakers encode the conceptually salient object of the sentence first, following the linearly order in which each concept needs to be uttered. On the other hand, prior work has shown that speakers do not always engage in linear planning (Griffin & Bock, 2000; Lee et al., 2013; Momma et al., 2016) and it is known that subjecthood holds a privileged status in other kinds of linguistic processes (Crawley & Stevenson, 1990; Gordon, Grosz, & Gilliom, 1993; Kaiser, 2011; *inter alia*). This suggests that the positional approach should not be taken as a given.

In addition to probing the tension between positional and functional processes in Experiment 1 (on English), in Experiment 2 we investigate at the production of declaratives and object *wh*-question in Mandarin Chinese to investigate effects of information focus – a third factor that could independently influence the process of linguistic encoding. Given that the *wh*-phrase in *wh*-questions is informationally focused, a possible concern is that eye-movement patterns during the production of object *wh*-questions could be influenced by these discourse-pragmatic properties, in addition to linear order and subjecthood. We address this possibility by looking at Mandarin, where declaratives and *wh*-questions have the same word order.

The results of Experiment 1 (English) showed that even though the articulation of a sentence is necessarily linear, speakers' *do not* encode sentences simply in accordance with the linear order in which the words are uttered. Evidence for this comes from speakers' eye-movements when they were encoding English object *wh*-questions: Even though the linear order of object *wh*-questions requires English speakers to say the *object* first, speakers nevertheless look to the *subject* during the window of encoding before looking to the object. The results of Experiment 2 (Mandarin) largely exclude the possibility that the eye-movement patterns in Experiment 1 were confounded by broader discourse-pragmatic effects (e.g. information focus) associated with *wh*-questions: In Experiment 2, we found no evidence that information focus could independently affect speakers' eye-movements during encoding.

As a whole, our results inform theories of linguistic encoding by providing real-time eye-tracking data showing that even in constructions where positional and functional processes are directly at odds with each other, we nevertheless observe effects of functional processes during the time period when linguistic encoding typically occurs. Thus, our eye-movement data – which allows us to very directly measure which elements speakers attend to at different points during the planning process – complements other studies which used speech-based measures (e.g. onset latencies) to investigate speech planning of other kinds of linguistic structures. Work by Lee et al. (2013), for instance, looked at the production of sentences with relative clause attachment ambiguities like '*Click on the fork of the king that's above the apple.*' and found evidence from speech onset latencies and word durations that speakers encoded their sentences following the syntactic structure, not the linear word order, of the utterance. Additionally, work by Momma et al. (2016) used the picture-word interference paradigm to investigate the production of verb-final sentences in Japanese. Using data from speech onset latencies and noun durations, this work also found evidence that linguistic encoding need not proceed strictly following the linear word order of the sentence.

Furthermore, our finding that object *wh*-questions elicit early looks to the *subject* despite the object being articulated first contributes novel information about how an under-researched but communicatively crucial construction – *wh*-questions – is produced and also extends prior work in several ways. First, our results are striking in that they speak to the centrality of hierarchical grammatical systems in language production processes: Even when the linear word order of object *wh*-questions effectively 'stacks the deck' against a functionally-driven approach to encoding, speakers nevertheless plan the linearly down-stream subject first. What's more, in any real-time production scenario, the benefits of constructing a more extensive syntactic frame ultimately trade off with communicative efficiency (i.e. more down-stream syntactic planning may lead to a more delayed speech onset). Our results also show, then, that in spite of this trade-off, speakers prioritize processes related to constructing this syntactic frame, even if this may run contrary to more immediate communicative demands.

In addition, our results speak to the nature of the *interaction* between functional and positional processes, and by extension, how conceptual versus linguistic representations are coordinated in real time. Specifically, we show that linguistic encoding is neither purely positionally- nor purely structurally-driven; rather, these processes are tightly linked temporally. In particular, although we find that subjecthood assignment – and by extension, functional processing – appears to be privileged during the linguistic encoding of English object *wh*-questions, as shown by speakers' initial looks to the subject, we also observe that the proportion of looks to the object increase very soon thereafter. This is clearly not predicted by a purely positionally-driven account. However, it is also not predicted by a purely functional account, which predicts that eye-movements during the encoding of object *wh*-questions should look just as they do with declaratives: Look to the subject should increase rapidly while looks to every other element in the utterance remain low until well after the earliest moments of encoding (i.e. until immediately before speech begins). This is not, however, what we observed in speakers' eye-movements when they are preparing object *wh*-questions: Instead, we find that although speakers look to the subject first, they quickly shift their gaze away from the subject to the linearly-initial object, presumably in preparation for encoding the object. Thus, linguistic encoding does not appear to be an 'all-or-nothing' process: *Both* positionally-driven and functionally-driven effects can be observed during the early moments encoding of object *wh*-questions.

Interestingly, prior work has suggested that earlier processes in production planning – namely, message formulation – are multi-factorial (Konopka & Kuchinsky, 2015; Konopka, 2012; Kuchinsky et al., 2011). That is, there is a hierarchy of factors that act in competition to influence the course of message formulation, which in turn, influences the course of linguistic encoding. The experiments reported in this paper suggest that the process of linguistic encoding also fits in a multiple-constraints framework: When speakers begin linguistically encoding their messages, multiple processes (both functional and positional) can play a role in determining what speakers encode first. In adjudicating between those processes, though, it appears that subjecthood assignment (a functional constraint) plays a more important role in determining the course of linguistic encoding than the order of the words in the sentence (a positional constraint). Thus, while encoding may be positionally-driven in many instances, whether speakers do ultimately adopt a highly linear approach may depend on the syntactic demands of the sentence.

Furthermore, it is worth emphasizing that our current understanding of the role of information focus on the time course of production planning is extremely limited: In fact, as far as we are aware, only Ganushchak et al. (2014) has done work in this area and only looked at effects of information focus in answers to questions. The results of our Experiment 2 did not detect any differences in Mandarin speakers' eye-movements as a function of informational focus. Even so, they do serve to clarify the results of Experiment 1 – namely, that the English results are unlikely to have been confounded by information focus. Moreover, especially in the context of recent work by Ganushchak et al. (2014), our results suggest that, if information focus affects production, its effects are presumably relatively small compared to effects of linear word order and subjecthood. Crucially, whether due to the sensitivity of our methods or to the nature of informational focus in production, our Mandarin results highlight the importance of future production work on the topic of information focus, and information structure more generally.

Finally, the observation that linguistic planning can be modulated by various factors raises the question of whether the privileged status of the subject (observed in our studies) is also prone to modulation. Indeed, other work has shown that under some instances, encoding proceeds linearly, driven by the immediate conceptual accessibility of a single lexical unit (i.e. a word; Gleitman et al., 2007; Brown-Schmidt & Konopka, 2008). Given the strength of the subject-preference that we observed – namely, that speakers begin linguistic encoding with the subject even when it is in a linearly disadvantaged position – our work opens the door for future work investigating whether and in what contexts/situations speakers would deprioritize encoding of the subject.

Conclusion

We conducted two visual-world eye-tracking experiments that investigate whether the transition from abstract conceptual representations to highly structured linguistic ones – known as linguistic encoding – begins with encoding of the subject (as predicted by functional accounts) or with encoding of the linearly initial element of the sentence (as predicted by positional accounts). In Experiment 1, we use object *wh*-questions ('*Which chefs did the nurses tickle?*') to pull apart effects of subjecthood from linear word order. We find evidence that (i) speakers prioritize encoding of the subject, even when it is not the linearly initial element in the sentence and (ii) functional and positional processes are tightly linked temporally. In Experiment 2, conducted in Mandarin Chinese, we investigated the potential influence of information focus on the starting point of linguistic encoding. Eye-movements do not show an independent effect of information focus on linguistic encoding. Our results provide the first real-time, cross-linguistic investigation of the production of *wh*-questions, which constitute a fundamental part of human communication but had not previously been explored using on-line methods.

Acknowledgments

We thank Pengchen Zhao, Shengjie Tang, and Zesheng Chen for their help with creating the stimuli and processing the data for the experiment in Mandarin. We are also grateful to Toben Mintz, Stefan Keine, Andrew Simpson, Rand Wilcox, Sarah Brown-Schmidt, Vic Ferreira, and the USC Language Processing Lab for their valuable insight and feedback on this work. We thank the audiences at the 23rd Architectures and Mechanisms for Language Processing Conference, the 1st California Meeting on Psycholinguistics, the 92nd Meeting of the Linguistic Society of America, and the 31st Annual CUNY Sentence Processing Conference, where earlier versions of some of this work were presented. Finally, we are very thankful for comments by the reviewers for their suggestions in improving this work. An earlier version of some of this work is included in the Proceedings of the 92nd Meeting of the Linguistic Society of America.

Conflicts of interest

None.

A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jml.2018.11.001.

References

Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53*(A), 741–764.
Altman, L. J. P., & Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Language and Cognitive Processes, 21*, 322–354.
Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language, 51*, 23–54.
Aoun, J., & Li, A. Y.-H. (1993). *Wh*-elements in situ: Syntax or LF? *Linguistic Inquiry, 24*, 199–238.
Arnold, J., Brown-Schmidt, S., & Trueswell, J. C. (2007). Children's use of gender and order-of- mention during pronoun comprehension. *Language and Cognitive Processes, 22*(527), 565.
Arnold, J., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language, 76*, 28–55.
Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R.* Cambridge: Cambridge University Press.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software, 67*, 1–48.
Birner, B. J., & Ward, G. (1998). *Information Status and Noncanonical Word Order in English.* Amsterdam: Johns Benjamins.
Bock, J. K. (1977). The effect of pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning and Verbal Behavior, 16*, 723–734.
Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575–586.
Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.
Bock, J. K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. Henderson, & F. Ferreira (Eds.). *The integration of language, vision, and action: Eye movements and the visual world* (pp. 249–278). New York: Psychology Press.
Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review, 99*, 150–171.
Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language, 48*, 653–685.
Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer. Retrieved from http://www.praat.org/.
Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition, 109*, 274–280.
Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language, 54*, 592–609.
Cheng, L. L -S. (1991). On the typology of wh-questions (Doctoral dissertation).
Chomsky, N. (1977). On wh-movement. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.). *Formal syntax* (pp. 71–132). New York: Academic Press.
Christensen, R. H. B. (2015). Ordinal - Regression models for ordinal data. R package version 2015.6-28. http://www.cran.r-project.org/package=ordinal/.
Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sen- tence production in a free word order language (Odawa). *Cognition, 98*, 105–135.
Crawley, R. J., & Stevenson, R. J. (1990). Reference in single sentences and in texts. *Journal of Psycholinguistic Research, 19*, 191–210.
Dryer, M. S. (2008). Order of subject, object, and verb. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.). *The world atlas of language structures online* < *http://wals.info* > . Munich: Max Planck Digital Library.
Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57–84.
Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language, 35*, 724–755.
Ferreira, V. S. (2010). Language production. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 834–844.
Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research, 32*, 669–692.
Frazier, L., & Flores d'Arcais, G. B. (1989). Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28*, 331–344.
Ganushchak, L. Y., Konopka, A. E., & Chen, Y. (2014). What the eyes say about planning of focused referents during sentence formulation: A cross-linguistic investigation. *Frontiers in Psychology, 5*.
Ganushchak, L. Y., Konopka, A. E., & Chen, Y. (2017). Accessibility of referent information influences sentence planning: An eye-tracking study. *Frontiers in Psychology,* 8.
Gelman, A., & Park, D. K. (2008). Splitting a predictor at the upper quarter or third and

the lower quarter or third. *The American Statistician, 62*, 1–8.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.). *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). Cambridge, MA: MIT Press.

Glaser, W. R., & Glaser, M. O. (1989). Context effects on stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*, 13–42.

Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the give- and-take between event apprehension and utterance formulation. *Journal of Memory and Language, 57*, 544–569.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17*, 311–347.

Greenberg, J. H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.). *Universals of language* (pp. 73–113). London: MIT Press.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*, B1–B14.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin and Review, 10*, 603–609.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of spoken language production. In M. J. Traxler, & M. A. Gernsbacher (Eds.). *Handbook of psycholinguistics* (pp. 21–59). (2nd ed.). London: Elsevier.

Hoff-Ginsberg, E. (1985). Some contributions of mothers' speech to their children's syntactic growth. *Journal of Child Language, 12*, 367–385.

Huang, C. -T. J. (1982). Logical relations in Chinese and the theory of grammar (Doctoral dissertation).

Hwang, H., & Kaiser, E. (2014). The role of the verb in grammatical function assignment in English and Korean. *Journal of Experimental Psychology, 40*, 1363–1376.

Iwasaki, N., Vinson, D. P., Vigliocco, G., Watanabe, M., & Arciuri, J. (2008). Naming action in Japanese: Effects of semantic similarity and grammatical class. *Language and Cognitive Processes, 23*, 889–930.

Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes, 26*, 1625–1666.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94*, 113–147.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition, 112*, 55–80.

Koenig, J. P., Mauner, G., & Bienvenue, B. (2003). Arguments for adjuncts. *Cognition, 89*, 67–103.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language, 66*, 143–162.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology, 73*, 1–40.

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.). *Oxford library of psychology. The Oxford handbook of language production* (pp. 3–20). New York: Oxford University Press.

Konopka, A. E., & Kuchinsky, S. E. (2015). How message similarity shapes the timecourse of sentence formulation. *Journal of Memory and Language, 84*, 1–23.

Kuchinsky, S. E. (2009). From seeing to saying: Perceiving, planning, producing (Doctoral Dis-sertation). Retrieved from ProQuest Dissertations and Theses.

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 748–756.

La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Experimental Psychology, 81*, 511–525.

Lee, E., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Journal of Memory and Language, 129*, 544–562.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 503–535.

Mahon, B. Z., Costa, A., Shapiro, K., & Caramazza, A. (2002). Grammatical class constraints on lexical selection in speech production: Functional and neuroanatomical considerations. Poster presented at the 2002 Science of Aphasia annual conference, Maratea, Italy. June 2002.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception and Psychophysics, 53*, 372–380.

McRae, K., Hare, M., Elman, J., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory and Cognition, 33*, 1174–1184.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*, 477–496.

Momma, S., Slevc, L. R., & Phillips, C. (2016). The timing of verb selection in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 813–824.

Myachykov, A., & Tomlin, R. S. (2008). Perceptual priming and structural choice in Russian sentence production. *Journal of Cognitive Science, 6*, 31–48.

Myachykov, A., Garrod, S., & Scheepers, C. (2010). Perceptual priming of structural choice during English and Finnish sentence production. In R. K. Mishra, & N. Srinivasan (Eds.). *Language and cognition: State of the art* (pp. 54–72). Munich: Lincom Europa.

Myachykov, A., Thompson, D., Scheepers, C., & Garrod, S. (2011). Visual attention and structural choice in sentence production across languages. *Language and Linguistic Compass, 5*, 95–107.

Myachykov, A., Thompson, D., Scheepers, C., & Garrod, S. (2012). Determinants of structural choice in visually situated sentence production. *Acta Psychologica, 141*, 304–315.

Neeleman, A., & Reinhart, T. (1997). Scrambling and the PF interface. In M. Butt, W. Gueder, & Butt (Eds.). *The projection of arguments* (pp. 309–353). Stanford: CSLI Publications.

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. (2015). Word order affects the timecourse of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience, 30*, 1187–1208.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language, 42*, 168–182.

R Core Team (2016). R: A language and environment for statisticalcomputing. R Foundation for Statistical Computing, Vienna, Austria. < http://www.R-project.org/ > .

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition, 47*, 59–87.

Ross, J. R. (1967). Constraints on Variables in Syntax (Doctoral Dissertation).

Rowe, M. L., Leech, K. A., & Cabrera, N. (2016). Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive Science, 41*, 162–179.

Rowland, C. F., Pine, J. M., Lieven, E. V., & Theakston, A. L. (2003). Determinants of acquisition order in wh-questions: Re-evaluating the role of caregiver speech. *Journal of Child Language, 30*, 609–635.

Sauppe, S., Norcliffe, E., Konopka, A. E., van Valin, R. D., & Levinson, S. (2013). Dependencies first: Eye-tracking evidence from sentence production in Tagalog. Paper presented at the 35th meeting of the cognitive science society, Berlin, Germany, July 2013.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture–word interference studies. *Journal of Memory and Language, 29*, 86–102.

Schriefers, H., Teruel, E., & Meinshausen, R. M. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language, 39*, 609–632.

Sekerina, I. A. (1997). The Syntax and Processing of Scrambling Constructions in Russian (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 1*, 227–245.

Sussman, R. (2006). Verb-instrument information during online-processing (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses.

Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: Evidence from eye-movements. *Language and Cognitive Processes, 18*, 143–163.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language, 65*, 318–330.

Tomlin, R. S. (1995). Focal attention, voice, and word order. In P. Dowing, & M. Noonan (Eds.). *Word order in discourse* (pp. 517–552). Amsterdam: John Benjamins.

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts, & E. Pederson (Eds.). *Language and conceptualization* (pp. 162–189). Cambridge: Cambridge University Press.

Ueno, M., & Kluender, R. (2009). On the processing of Japanese *wh*-questions: An ERP study. *Brain Research, 1290*, 63–90.

van Nice, K. Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture description. *Linguistics, 41*, 825–849.

Wagers, M. (2013). Memory mechanisms for wh-dependency formation and their implications for islandhood. In J. Sprouse, & N. Hornstein (Eds.). *Experimental syntax and island effects* (pp. 161–185). Cambridge: Cambridge University Press.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 423–440.

Wilcox, R. R. (2017). *Modern statistics for the social and behavioral sciences: A practical introduction* (2nd ed.). New York: Chapman & Hall/CRC Press.

Wilcox, R. R. & Rousselet, G.A (2017). A guide to robust statistical methods: Illustrations using R https://doi.org/10.6084/m9.figshare.5114275.

Wilcox, R. R., & Keselman, H. J. (2012). Modern regression methods that can substantially increase power and provide a more accurate understanding of associations. *European Journal of Personality, 26*, 165–174.

Xiang, M., Dillon, B., Wagers, M., Liu, F. Q., & Guo, T. M. (2014). Processing covert dependencies: An SAT study on Mandarin *wh*-in-situ questions. *Journal of East Asian Linguistics, 23*, 207–232.

Xiang, M., Wang, S.-P., & Cui, Y.-L. (2015). Constructing covert dependencies—The case of Mandarin *wh*-in-situ dependency. *Journal of Memory and Language, 84*, 139–166.